The maximum likelihood estimates are:

$$\tilde{\beta}_2 = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2}, \qquad \tilde{\beta}_1 = \overline{y} - \tilde{\beta}_2 \overline{x}, \qquad \tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \tilde{\beta}_1 - \tilde{\beta}_2 x_i)^2.$$

The MLE of $\sigma^2$ is divided by $n$, not $n-2$.

## 9.2 MLE: The Case of Multiple Regression Model I

1. Multivariate Normal Distribution: $X:\ n \times 1$ and $X \sim N(\mu, \Sigma)$

   The density function of $X$ is:

$$f(x) = (2\pi)^{n/2} |\Sigma|^{-1/2} \exp\Big(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\Big).$$

2. Regression model: $y = X\beta + u, \quad u \sim N(0, \sigma^2 I_n)$

Transformation of Variables from $u$ to $y$:

$$f_u(u) = (2\pi\sigma^2)^{-n/2} \exp\Big(-\frac{1}{2\sigma^2} u'u\Big)$$

$$\begin{aligned} f_y(y) &= f_u(y - X\beta) \left|\frac{\partial u}{\partial y'}\right| \\ &= (2\pi\sigma^2)^{-n/2} \exp\Big(-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\Big) \\ &= L(\theta; y, X), \end{aligned}$$

where $\theta = (\beta, \sigma^2)$, because of $\dfrac{\partial u}{\partial y'} = I_n$.

Therefore, the log-likelihood function is:

$$\log L(\theta; y, X) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta),$$

Note that $|\Sigma|^{-1/2} = |\sigma^2 I_n|^{-1/2} = \sigma^{-n/2}$.

3. $\max_{\theta} \quad \log L(\theta; y, X)$

(FOC) $\dfrac{\partial \log L(\theta; y, X)}{\partial \theta} = 0$

(SOC) $\dfrac{\partial^2 \log L(\theta; y, X)}{\partial \theta \partial \theta'}$ is a negative definite matrix.

We obtain MLE of $\beta$ and $\sigma^2$:

$$\tilde{\beta} = (X'X)^{-1}X'y, \qquad \tilde{\sigma}^2 = \frac{(y - X\tilde{\beta})'(y - X\tilde{\beta})}{n},$$

where $\tilde{\sigma}^2$ is divided by $n$, not $n - k$.

4. Fisher's information matrix is:

$$I(\theta) = -\mathrm{E}\Big(\frac{\partial^2 \log L(\theta; y, X)}{\partial \theta \partial \theta'}\Big)$$

The inverse of the information matrix, $I(\theta)^{-1}$, provides a lower bound of the variance - covariance matrix for unbiased estimators of $\theta$ .

$$I(\theta)^{-1} = \begin{pmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{pmatrix}$$

For large $n$, we approximately obtain: $\begin{pmatrix} \tilde{\beta} \\ \tilde{\sigma}^2 \end{pmatrix} \sim N\left(\begin{pmatrix} \beta \\ \sigma^2 \end{pmatrix}, \begin{pmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{pmatrix}\right).$

## 9.3 MLE: The Case of Multiple Regression Model II

1. Regression model: $\quad y = X\beta + u, \quad u \sim N(0, \sigma^2\Omega)$

   Transformation of Variables from $u$ to $y$:

$$f_u(u) = (2\pi\sigma^2)^{-n/2}|\Omega|^{-1/2}\exp\Big(-\frac{1}{2\sigma^2}u'\Omega^{-1}u\Big)$$

$$f_y(y) = f_u(y - X\beta)\left|\frac{\partial u}{\partial y'}\right|$$

$$= (2\pi\sigma^2)^{-n/2}|\Omega|^{-1/2}\exp\Big(-\frac{1}{2\sigma^2}(y - X\beta)'\Omega^{-1}(y - X\beta)\Big)$$

$$= L(\theta; y, X),$$

where $\theta = (\beta, \sigma^2)$, because of $\dfrac{\partial u}{\partial y'} = I_n$.

The log-likelihood function is:

$$\log L(\theta; y, X) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log|\Omega| - \frac{1}{2\sigma^2}(y - X\beta)'\Omega^{-1}(y - X\beta),$$

where $\theta = (\beta, \sigma^2)$.

2. $\max_{\theta} \log L(\theta; y, X)$

(FOC) $\dfrac{\partial \log L(\theta; y, X)}{\partial \theta} = 0$

(SOC) $\dfrac{\partial^2 \log L(\theta; y, X)}{\partial \theta \partial \theta'}$ is a negative definite matrix.

Then, we obtain MLE of $\beta$ and $\sigma^2$:

$$\tilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y, \qquad \tilde{\sigma}^2 = \frac{(y - X\tilde{\beta})'\Omega^{-1}(y - X\tilde{\beta})}{n}$$

3. Fisher's information matrix is defined as:

$$I(\theta) = -\mathrm{E}\Big(\frac{\partial^2 \log L(\theta; y, X)}{\partial\theta\partial\theta'}\Big)$$

The inverse of the information matrix, $I(\theta)^{-1}$, provides a lower bound of the variance - covariance matrix for unbiased estimators of $\theta$, which is given by:

$$I(\theta)^{-1} = \begin{pmatrix} \sigma^2(X'\Omega^{-1}X)^{-1} & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{pmatrix}$$

## 9.4 MLE: AR(1) Model

The $p$th-order Autoregressive Model, i.e., AR($p$) Model ($p$ 次の自己回帰モデル):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t$$

AR(1) Model: $t = 2, 3, \cdots, n,$

$$y_t = \phi_1 y_{t-1} + u_t, \quad u_t \sim N(0, \sigma^2)$$

where $|\phi_1| < 1$ is assumed for now.

To obtain the joint density function of $y_1, y_2, \cdots, y_n$, $f(y_n, y_{n-1}, \cdots, y_1)$ is decomposed as follows:

$$
\begin{aligned}
& f(y_n, y_{n-1}, \cdots, y_1) \\
&\quad = f(y_n|y_{n-1}, \cdots, y_1) f(y_{n-1}, y_{n-2}, \cdots, y_1) \\
&\quad = f(y_n|y_{n-1}, \cdots, y_1) f(y_{n-1}|y_{n-2}, \cdots, y_1) f(y_{n-2}, y_{n-3}, \cdots, y_1) \\
&\qquad\quad \cdots \\
&\quad = f(y_n|y_{n-1}, \cdots, y_1) f(y_{n-1}|y_{n-2}, \cdots, y_1) f(y_{n-2}, y_{n-3}, \cdots, y_1) \ \cdots \ f(y_2|y_1) f(y_1) \\
&\quad = f(y_1) \prod_{t=2}^{n} f(y_t|y_{t-1}, \cdots, y_1).
\end{aligned}
$$

Note that Bayes theorem is applied and repeated.

That is, $P(A \cap B) = P(A|B)P(B)$ for two events $A$ and $B$.

We say that the joint distribution (or the likelihood function) is represented in the **innovation form**.

From $y_t = \phi_1 y_{t-1} + u_t$, we can obtain:

$$\mathrm{E}(y_t|y_{t-1}, \cdots, y_1) = \phi_1 y_{t-1}, \quad \text{and} \quad \mathrm{V}(y_t|y_{t-1}, \cdots, y_1) = \sigma^2.$$

Therefore, the conditional distribution $f(y_t|y_{t-1}, \cdots, y_1)$ is:

$$f(y_t|y_{t-1}, \cdots, y_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_t - \phi_1 y_{t-1})^2\right).$$

To obtain the unconditional distribution $f(y_t)$, $y_t$ is rewritten as follows:

$$
\begin{aligned}
y_t &= \phi_1 y_{t-1} + u_t \\
&= \phi_1^2 y_{t-2} + u_t + \phi_1 u_{t-1} \\
&\ \ \vdots \\
&= \phi_1^\tau y_{t-\tau} + u_t + \phi_1 u_{t-1} + \cdots + \phi_1^{\tau-1} u_{t-\tau+1} \\
&\ \ \vdots \\
&= u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \cdots, \quad \text{when } \tau \text{ goes to infinity under the condition } |\phi_1| < 1.
\end{aligned}
$$